Image and Vision Computing 58 (2017) 38–46

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Pose-specific non-linear mappings in feature space towards multiview facial expression recognition☆

Mahdi Jampour*, Vincent Lepetit, Thomas Mauthner, Horst Bischof

Graz University of Technology, Graz, Austria

ARTICLE INFO

Article history:
Received 28 September 2015
Received in revised form 31 March 2016
Accepted 5 May 2016
Available online 13 May 2016

Keywords:
Non-frontal facial expression recognition
Sparse coding
Non-linear transformation
Robust arbitrary view facial expression recognition

ABSTRACT

We introduce a novel approach to recognizing facial expressions over a large range of head poses. Like previous approaches, we map the features extracted from the input image to the corresponding features of the face with the same facial expression but seen in a frontal view. This allows us to collect all training data into a common referential and therefore benefit from more data to learn to recognize the expressions. However, by contrast with such previous work, our mapping depends on the pose of the input image: We first estimate the pose of the head in the input image, and then apply the mapping specifically learned for this pose. The features after mapping are therefore much more reliable for recognition purposes. In addition, we introduce a non-linear form for the mapping of the features, and we show that it is robust to occasional mistakes made by the pose estimation stage. We evaluate our approach with extensive experiments on two protocols of the BU3DFE and Multi-PIE datasets, and show that it outperforms the state-of-the-art on both datasets.

1. Introduction

With the ever growing importance of Human–Computer Interfaces, facial expression recognition (FER) is one of the important challenges of computer vision. Even if recognition in frontal views, either based on appearance or geometry already performs very well [1,2,3,4,5], having a frontal view is an unrealistic assumption for real-world applications, and multiview facial expression recognition (MFER) is still very challenging as facial features important for recognition are likely to be hidden.

To date, the most successful methods [6,7,8] map facial features extracted from non-frontal views to the corresponding features in the frontal view: by mapping all the available training data to a common referential one can generalize better. However, Ref. [6] used the same mapping regardless of the pose of the head; Ref. [7] proposed a complex method that relies on a time-consuming optimization process.

We therefore propose to learn several mappings, each specific to a pose from a discrete set of possible poses of the face. To know which mapping to use for a new input image, we simply rely on another classifier to predict the pose of the face. Since the mappings are adapted to the pose of the input face, this approach yields significantly better results than using a single mapping. Fig. 1 illustrates our claim that local mappings can provide more reasonable results than a global one. We explore two different forms to perform the mappings in feature space: We first consider a simple linear mapping, and we introduce non-linear mappings based on Taylor expansion.

We evaluate this approach with extensive experiments on two protocols of the BU3DFE and Multi-PIE datasets. Our evaluation shows that simple linear transformations for the mappings are enough for our approach to outperform the state-of-the-art on both datasets. When non-linear mappings are used, we improve the results even further.

This paper is an extension of our previous work [9], where we introduced pose specific linear mappings. Here we introduce non-linear mappings, and provide an extensive evaluation, and an in-depth discussion.

In the remainder of the paper, we first discuss related work. We then explain how we predict the pose of the faces in the input images. After that, we describe the different mappings and the facial expression recognition, and finally we provide extensive evaluation of our approach.

2. Related work

Facial expression recognition has many exciting and various applications, including Human–Computer Interaction (HCI), psychology, games, children education, etc., and the literature is

☆ This paper has been recommended for acceptance by Vitomir Štruc.

* Corresponding author.
E-mail addresses: jampour@icg.tugraz.at (M. Jampour), lepetit@icg.tugraz.at (V. Lepetit), mauthner@icg.tugraz.at (T. Mauthner), bischof@icg.tugraz.at (H. Bischof).

http://dx.doi.org/10.1016/j.imavis.2016.05.002
0262-8856/

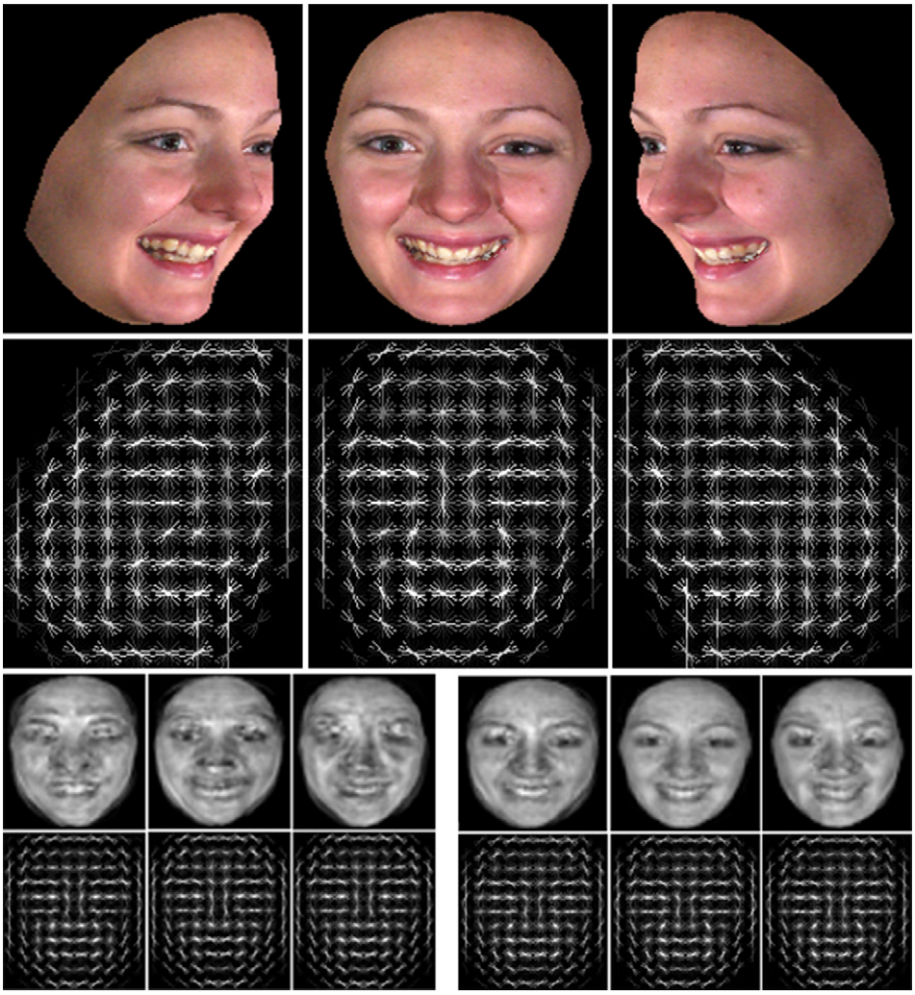

Fig. 1. Comparison between global and pairwise transformations: (a) three samples from different viewpoints and their HOG features; their reconstructions using (b) a global mapping and (c) a pairwise mapping. Note that the bitmap images are provided for visualization only. We use a concatenation of HOG and LBP features instead of raw features for recognition.

very broad. Some approaches explicitly consider the main facial components, mouths, eyes, etc. and extract them from the input images. Most of the geometric approaches use Facial Action Units (AUs) which are part of Facial Action Coding Systems (FACS) introduced by Ekman et al. [10] to recognize expressions. AUs are observable components of facial movement that acted by a group of facial muscles [11,12,13,14]. Other approaches [15,16,17,18,19] rely only on texture information and use local descriptors such as SIFT, Gabor, HOG, Pyramid HOG (PHOG), LBP, Pyramid LBP (PLBP). Some hybrid methods exploit both geometric and texture information [6,20,21,22].

Recently, researchers turned to the multiview facial expression recognition problem, where the face is not necessary frontal. This problem is of course much more challenging than recognizing facial expression from frontal, as the perspective can deform the expressions, or even hide some features.

One of the first attempts for non-frontal facial expression recognition was Ref. [23], which proposed to consider the 2D facial feature displacements of 38 facial landmarks as features to perform the recognition and they investigated various classifiers on the BU3DFE dataset. Rudovic et al. [24] proposed a mapping model between the facial points from non-frontal views to a frontal referential. They used Coupled Scaled Gaussian Process Regression (CSGPR) for the mapping, and multiclass LDA for estimating the head poses.

Other approaches rely on appearance only. For instance, Zheng [7] proposed a Group Sparse Reduced-Rank Regression based method (GSRRR). Sparse SIFT descriptors and LBP features are extracted. Feature vectors are synthesized using kernel reduced-rank regression to obtain feature vectors corresponding to different facial views. The facial expression category is finally obtained by solving GSRRR optimization problem. This approach obtains very good results, however it is very computationally expensive. The same author proposed a discriminant analysis theory (BDA/GMM) [15] which optimizes the upper bound of the Bayes error derived by Gaussian mixture model but it only outperformed the baseline.

Hesse et al. [16] evaluated various descriptors such as SIFT, LBP and DCT extracted around facial landmarks and classify then using ensemble SVM. They showed that DCT based features yield better performance than SIFT or LBP. Another approach proposed by Moore and Bowden [25] first estimates the pose orientation directly from the image and then applies a pose-dependent classifier to recognize the facial expressions. This method is simple and fast as it is based on the LBP features. However, it is sensitive to occlusion, and more importantly to the low number of training data for each specific head pose.

Huang et al. [18] proposed a discriminative framework based on multi-set canonical correlation analysis (MCCA) and proposed a multiview model theorem for facial expression recognition with